

This document is an English translation of the abstract of Korean Patent Application Publication No. 2001-0066401.

#### Abstract

The present invention relates to a method of compressing and decompressing Korean alphabet data.

A method of compressing according to the present invention includes: a step of compressing file by classifying types of file to be compressed, generating index table/housing table for compression code by taking character stream's file to be compressed after initializing the dictionary and generating index table in memory; and a step of remembering index number for compression code, outputting coded data from code stream, checking whether there exist more data to be compressed and outputting code word + prefix code of the code stream.

On the other hand, a method of decompressing according to the present invention includes: a step of compressing file by classifying type of file if file of data to be decompressed is inputted, outputting character word of character stream by taking original code of code stream after initializing the dictionary, substituting character word for prefix word and checking whether next character word exists in dictionary; a step of substituting code word taken as character stream for prefix after outputting the code word and adding new data in the dictionary by taking the new data from the code word, if next character word exists in the dictionary; and a step of substituting code stream for prefix after taking prefix by reading and checking whether there exists other data to be decompressed in code stream by outputting prefix + code word into character stream and by adding output value into the dictionary.

Thus, it is effective to use the present invention for compressing or decompressing all data using Korean alphabet.

# (19) 대한민국특허청(KR) (12) 공개특허공보(A)

(51) Int. Cl. <sup>7</sup> H04N 7/025	(11) 공개번호 (43) 공개일자	특2001-0066401 2001년07월11일
(21) 출원번호	10-1999-0068112	
(22) 출원일자	1999년12월31일	
(71) 출원인	(주) 한국인프라	
(72) 발명자	서울 강남구 도곡2동 435-16 김기영	
(74) 대리인	서울특별시송파구송파동한양아파트31동809호 류명현, 안종철, 이금호	

심사청구 : 있음

## (54) 한글 데이터의 압축 및 복원방법

### 요약

본 발명은 한글 데이터의 압축 및 복원방법에 관한 것으로, 압축시킬 파일을 입력시키면, 파일의 형태를 분류하여 압축하고 그 dictionary를 초기화한 후 캐릭터 스트림의 압축할 파일을 가져와 압축 코드에 대한 인덱스 테이블/하우징 테이블을 생성하고 메모리에 인덱스 테이블을 생성하는 단계와,

압축 코드에 대한 인덱스 번호를 기억시키고 코드화된 자료를 코드 스트림에서 출력하면서 압축할 자료가 더 있는가를 확인하면서 코드워드 + 코드 스트림의 프리픽스(prefix) 코드를 출력하는 단계들에 의해 압축을 수행하도록 하는 한편,

사용자가 복원하고자 하는 데이터의 파일을 입력시키면, 파일의 형태를 분류하여 파일을 압축하고 그 dictionary를 초기화한 후 코드 스트림의 처음 코드를 가져와 캐릭터 스트림의 캐릭터 워드를 출력하고 프리픽스 워드를 캐릭터 워드로 대체하도록 하고 다음에 위치하는 캐릭터 워드를 가져와 캐릭터 워드가 dictionary에 있는가를 확인하는 단계와,

dictionary에 있는 경우에는 캐릭터 스트림으로 가져온 코드 워드를 출력한 후 프리픽스로 대체하도록 하고 코드워드에서 새로운 데이터를 가져와 프리픽스 새 데이터를 dictionary에 추가하는 단계와;

상기의 dictionary에 없는 경우에는 프리픽스를 읽어서 가져온 후 코드 스트림을 프리픽스로 대체하도록 하고 프리픽스 + 코드 워드를 캐릭터 스트림에 출력하면서 dictionary에 출력값을 추가하여 코드 스트림에 다른 해석할 자료가 더 있는가를 확인하는 과정들에 의해 복원을 수행하도록 함으로써 한글을 사용하는 모든 데이터의 압축이나 복원에 효율적으로 이용할 수 있도록 한 것이다.

### 대표도

### 도1

### 명세서

#### 도면의 간단한 설명

도 1은 본 발명의 데이터를 압축하는 과정을 나타낸 플로우 차트.

도 2는 본 발명의 데이터를 복원하는 과정을 나타낸 플로우 차트.

도 3은 본 발명 자동 풀림 편집기의 메뉴를 나타낸 개략도.

도 4는 본 발명 압축파일 분할의 메뉴를 나타낸 개략도.

도 5는 본 발명 Zip파일 검사/치료의 메뉴를 나타낸 개략도.

#### 발명의 상세한 설명

##### 발명의 목적

##### 발명이 속하는 기술 및 그 분야의 종래기술

본 발명은 한글 데이터의 압축 및 복원방법에 관한 것으로, 특히 기존의 데이터를 압축하기 위한 프로그램들과의 호환성이 뛰어나며 압축과 복원 뿐아니라, 압축된 파일을 직접 E-mail로 보내거나 분할하여 압축시킬 수 있으며, 손상된 파일의 검색 및 치료 등의 기능을 갖도록 함으로써 한글을 사용하는 모든 데

이터의 압축이나 복원에 효율적으로 이용할 수 있도록 한 한글 데이터의 압축 및 복원방법에 관한 것이다.

일반적으로 컴퓨터를 사용하는 가정이나 직장이 많아지면서 거의 모든 정보를 컴퓨터용 파일로 하여 저장하는 경향이 많아지고 있음은 이미 주지된 사실이다.

그리고 이러한 컴퓨터 파일로 저장하는 것은 '0'이나 '1'의 데이터를 조합한 디지털 신호로 기록된 상태를 저장하는 것이므로 문자로만 이루어진 정보의 데이터를 저장할 경우에는 그 용량이 별로 크지 않아 디지털 신호의 조합으로 이루어진 많은 용량의 파일을 저장할 수 있지만, 오디오 신호 또는 그림이나 색채의 조합에 의한 디지털 정보인 경우에는 그 용량이 매우 커져서 하나의 저장용 디스켓을 사용할 수 없음은 물론, 웬만한 용량의 컴퓨터의 잔여 저장영역에도 저장할 수 없는 경우가 허다한 지경이다.

그러므로 일반 데이터를 압축하거나 복원할 수 있도록 한 윈Zip(Winzip)이 제안되어 사용하도록 하였다.

#### 발명이 이루고자하는 기술적 과제

그러나 상기와 같은 종래의 데이터의 압축 및 복원 방법에 의하여서는 이는 zip파일만 지원하면서 압축시킬 수 있으며, zip이외의 포맷을 복원하고자 할 경우에는 다른 포맷을 실행할 프로그램이 항상 존재하여야만 가능하고, 영문의 메뉴에만 사용할 수 있으며, 압축된 상태의 파일을 형식 변환할 수 없고, 압축된 상태에서 복원하면서 파일을 정상적으로 설치한 후에만 실행이 가능하므로 국내의 많은 컴퓨터를 사용하는 사용자들이 효율적으로 이용할 수 없는 등의 단점이 있었다.

따라서 본 발명은 기존의 데이터를 압축하기 위한 프로그램들과의 호환성이 뛰어나며 압축과 복원 뿐 아니라, 압축된 파일을 직접 E-mail로 보내거나 분할하여 압축시킬 수 있으며, 손상된 파일의 검색 및 치료 등의 기능을 갖도록 함으로써 한글을 사용하는 모든 데이터의 압축이나 복원에 효율적으로 이용할 수 있도록 한 한글 데이터의 압축 및 복원방법을 제공하는 것을 그 목적으로 한다.

이와 같은 목적을 달성하기 위한 본 발명은 먼저 압축 및 복원을 위한 프로그램이 압축된 CD 등을 컴퓨터의 CD 롬 드라이버에 넣은 후 프로그램을 읽도록 명령하면, 컴퓨터에서 CD 롬 드라이버를 가동하면서 픽업 모듈을 통하여 저장된 데이터를 읽는 단계와,

상기의 CD 롬 드라이버를 통하여 CD의 데이터를 읽은 초기화면을 디스플레이하면서 사용자가 프로그램을 설치할 것인가의 여부를 묻는 단계와,

사용자가 프로그램을 설치하기 위해 설치의 버튼을 클릭하면 CD에 압축되어 저장된 프로그램을 읽어서 복원한 후 사용자가 지정한 번지에 저장하는 단계와,

그 상태에서 사용자가 압축시키고자 하는 데이터의 파일을 입력시키면, 파일의 형태를 분류하여 형태별로 구분하는 단계와,

상기의 구분된 형태들의 파일을 압축하는 단계와,

상기의 압축된 파일의 dictionary를 초기화한 후 캐릭터 스트림의 압축할 파일을 읽어서 가져오는 단계와,

압축 코드에 대한 인덱스 테이블/하우징 테이블을 생성한 후 메모리에 인덱스 테이블을 생성하는 단계와,

압축 코드에 대한 인덱스 번호를 기억시킨 후 코드화된 자료를 코드 스트림에서 출력하는 단계와,

캐릭터 스트림에 압축할 자료가 있는가를 확인하면서 dictionary를 초기화하는 단계부터 다시 수행하는 단계와,

캐릭터 스트림에 압축할 자료가 더 이상 없으면, 코드워드 + 코드 스트림의 프리픽스(prefix) 코드를 출력한 후 종료하는 단계들에 의해 압축을 수행하도록 하는 한편,

사용자가 복원하고자 하는 데이터의 파일을 입력시키면, 파일의 형태를 분류하여 형태별로 구분하는 단계와,

상기의 구분된 형태들의 파일을 압축하는 단계와,

상기의 압축된 파일의 dictionary를 초기화한 후 코드 스트림의 처음 코드를 읽어서 가져오는 단계와,

캐릭터 스트림의 캐릭터 워드를 출력하고 프리픽스 워드를 캐릭터 워드로 대체하는 단계와,

코드 스트림의 다음에 위치하는 캐릭터 워드를 읽어서 가져온 상태에서 가져온 캐릭터 워드가 dictionary에 있는가를 확인하는 단계와,

dictionary에 있는 경우에는 캐릭터 스트림으로 가져온 코드 워드를 출력한 후 프리픽스로 대체하는 단계와,

코드워드에서 새로운 데이터를 읽어서 가져온 후 프리픽스 새 데이터를 dictionary에 추가하는 단계와,

상기의 dictionary에 없는 경우에는 프리픽스를 읽어서 가져온 후 코드 스트림을 프리픽스로 대체하는 단계와,

프리픽스 + 코드 워드를 캐릭터 스트림에 출력하면서 dictionary에 출력값을 추가한 후 코드 스트림에 해석할 또 다른 해석할 자료가 더 있는가를 확인하는 과정들에 의해 복원을 수행하도록 함으로써 한글을 사용하는 모든 데이터의 압축이나 복원에 효율적으로 이용할 수 있도록 한 것이다.

### 발명의 구성 및 작용

이하 본 발명을 첨부 도면에 의거 상세히 기술하면 다음과 같다.

본 발명은 하드웨어적인 구성보다는 파일을 저장하면서 데이터를 압축하거나 재생하는 과정에 그 특징이 있는 것으로서,

먼저 압축 및 복원을 위한 프로그램이 압축된 CD 등을 컴퓨터의 CD 롬 드라이버에 넣은 후(단계 1) 프로그램을 읽도록 명령하면(단계 2), 컴퓨터에서 CD 롬 드라이버를 가동하면서 픽업 모듈을 통하여 저장된 데이터를 읽는다(단계 3).

상기의 CD 롬 드라이버를 통하여 CD의 데이터를 읽은 초기화면을 디스플레이하면서 사용자가 프로그램을 설치할 것인가의 여부를 묻고(단계 4), 사용자가 프로그램을 설치하기 위해 설치의 버튼을 클릭하면(단계 5), CD에 압축되어 저장된 프로그램을 읽어서 복원한 후 사용자가 지정한 번지에 저장한다(단계 6).

상기와 같이 데이터의 압축이나 복원을 위한 프로그램을 저장한 상태에서 사용자가 압축시키고자 하는 데이터의 파일을 입력시키면(단계 7), 압축할 파일이 형태를 구분할 수 있는가의 여부를 판단하여(단계 8) 가능하지 않으면 에러를 표시한 후 종료하고(단계 19), 파일의 구분이 가능하면 파일의 형태를 분류하여 ARJ, ARC, GZIP, LHA, LZH, PAK, RAR, TAR, ZIP, ZOO 등의 형태별로 구분한다(단계 9).

상기의 구분된 형태들의 파일을 'LZW'의 형태로 압축하고(단계 10), 상기의 압축된 파일의 dictionary를 초기화한 후(단계 11) 캐릭터 스트림의 압축할 파일을 읽어서 가져온다(단계 12).

압축 코드에 대한 인덱스 테이블/하우징 테이블을 생성한 후(단계 13), 메모리에 인덱스 테이블을 생성한다(단계 14).

압축 코드에 대한 인덱스 번호를 기억시킨 후(단계 15), 코드화된 자료를 코드 스트림에서 출력한다(단계 16).

캐릭터 스트림에 압축할 자료가 더 있는가를 확인하여(단계 17), 있으면 dictionary를 초기화하는 단계부터 다시 수행하도록 하는 한편, 캐릭터 스트림에 압축할 자료가 더 이상 없으면, 코드워드 + 코드 스트림의 프리픽스(prefix) 코드를 출력한 후 종료하는 단계들에 의해 압축을 수행하도록 한다(단계 18).

상기의 압축파일을 자동실행 파일(.exe)로 만들 수 있으며 이는 도 3에 도시한 것과 같은 자동폴름 편집의 메뉴를 선택하여(단계 20) Zip파일과 Exe 위치 및 표시제목을 입력시킴으로 가능하게 되고, setup.exe 같은 파일을 생성할 때 유용하게 쓰일 수 있게 된다(단계 21).

그리고 압축한 파일을 하나의 디스켓에 다 저장하지 못할 경우에는 도 4에 도시한 것과 같은 압축파일 분할의 메뉴를 선택하여(단계 22) 그 크기를 임의로 설정하면서 분할하여 저장할 수 있도록 한다(단계 23).

그리고 도 5에 도시한 것과 같은 Zip파일 검사/치료의 메뉴를 선택하면(단계 24) 압축파일을 검사하고 파일이 손상되었을 때 파일을 복구하고 새로운 이름으로 저장할 수 있다(단계 25).

한편, 상기의 압축과정에 의해 압축된 데이터를 사용자가 복원하고자 할 경우에는, 데이터의 파일을 입력시키면(단계 31), 압축할 파일이 형태를 구분할 수 있는가의 여부를 판단하여(단계 32) 가능하지 않으면 에러를 표시한 후 종료하는 한편(단계 33), 파일의 구분이 가능하면 파일의 형태를 분류하여 ARJ, ARC, GZIP, LHA, LZH, PAK, RAR, TAR, ZIP, ZOO 등의 형태별로 구분한다(단계 34).

상기의 구분된 형태들의 파일을 'LZW'와 같은 형태로 압축하고(단계 35), 상기의 압축된 파일의 dictionary를 초기화한 후(단계 36) 코드 스트림의 처음 코드를 읽어서 가져오도록 한다(단계 37).

캐릭터 스트림의 캐릭터 워드를 출력하고(단계 38) 프리픽스 워드를 캐릭터 워드로 대체한다(단계 39).

코드 스트림의 다음에 위치하는 캐릭터 워드를 읽어서 가져온 상태에서(단계 40) 가져온 캐릭터 워드가 dictionary에 있는가를 확인한다(단계 41).

dictionary에 있는 경우에는, 캐릭터 스트림으로 가져온 코드 워드를 출력한 후(단계 42) 프리픽스로 대체한다(단계 43).

코드워드에서 새로운 데이터를 읽어서 가져온 후(단계 44) 프리픽스 새 데이터를 dictionary에 추가한다(단계 45).

상기의 단계 41에서 dictionary에 없는 경우에는, 프리픽스를 읽어서 가져온 후(단계 46) 코드 스트림을 프리픽스로 대체한다(단계 47).

프리픽스 + 코드 워드를 캐릭터 스트림에 출력하면서 dictionary에 출력값을 추가한 후(단계 48) 코드 스트림에 해석할 또 다른 해석할 자료가 더 있는가를 확인하는 과정들에 의해 복원을 수행하도록 한다(단계 49).

### 발명의 효과

이와 같은 본 발명의 한글 데이터의 압축 및 복원방법에 의하여서는 압축시킬 파일을 입력시키면, 파일의 형태를 분류하여 압축하고 그 dictionary를 초기화한 후 캐릭터 스트림의 압축할 파일을 가져와 압축 코드에 대한 인덱스 테이블/하우징 테이블을 생성하고 메모리에 인덱스 테이블을 생성하고,

압축 코드에 대한 인덱스 번호를 기억시키고 코드화된 자료를 코드 스트림에서 출력하면서 압축할 자료가 더 있는가를 확인하면서 코드워드 + 코드 스트림의 프리픽스 코드를 출력하는 단계들에 의해 압축을 수행하도록 하는 한편,

복원하고자 하는 데이터의 파일을 입력시키면, 파일의 형태를 분류하여 파일을 압축하고 그 dictionary를 초기화한 후 코드 스트림의 처음 코드에 대해 캐릭터 스트림의 캐릭터 워드를 출력하고 프리픽스 워드를 캐릭터 워드로 대체하고 다음의 캐릭터 워드를 가져와 캐릭터 워드가 dictionary에 있는가를 확인하는 단계와,

dictionary에 있는 경우에는 캐릭터 스트림으로 가져온 코드 워드를 출력한 후 프리픽스로 대체하도록 하고 코드워드에서 새로운 데이터를 가져와 프리픽스 새 데이터를 dictionary에 추가하는 단계와,

상기의 dictionary에 없는 경우에는 프리픽스를 읽어서 가져온 후 코드 스트림을 프리픽스로 대체하도록 하고 프리픽스 + 코드 워드를 캐릭터 스트림에 출력하면서 dictionary에 출력값을 추가하여 코드 스트림에 다른 해석할 자료가 더 있는가를 확인하는 과정들에 의해 복원을 수행하도록 함으로써 한글을 사용하는 모든 데이터의 압축이나 복원에 효율적으로 이용할 수 있도록 한 것이다.

#### (57) 청구의 범위

##### 청구항 1

압축 및 복원을 위한 프로그램이 압축된 CD를 컴퓨터에 저장한 상태에서

사용자가 압축시키고자 하는 데이터의 파일을 입력시키면, 파일의 형태를 분류하여 형태별로 구분하는 단계와,

상기의 구분된 형태들의 파일을 압축하는 단계와,

상기의 압축된 파일의 dictionary를 초기화한 후 캐릭터 스트림의 압축할 파일을 읽어서 가져오는 단계와,

압축 코드에 대한 인덱스 테이블/하우징 테이블을 생성한 후 메모리에 인덱스 테이블을 생성하는 단계와,

압축 코드에 대한 인덱스 번호를 기억시킨 후 코드화된 자료를 코드 스트림에서 출력하는 단계와,

캐릭터 스트림에 압축할 자료가 있는가를 확인하면서 dictionary를 초기화하는 단계부터 다시 수행하는 단계와,

캐릭터 스트림에 압축할 자료가 더 이상 없으면, 코드워드 + 코드 스트림의 프리픽스 코드를 출력한 후 종료하는 단계들에 의해 압축을 수행하도록 한 한글 데이터의 압축방법.

##### 청구항 2

사용자가 복원하고자 하는 데이터의 파일을 입력시키면, 파일의 형태를 분류하여 형태별로 구분하는 단계와,

상기의 구분된 형태들의 파일을 압축하는 단계와,

상기의 압축된 파일의 dictionary를 초기화한 후 코드 스트림의 처음 코드를 읽어서 가져오는 단계와,

캐릭터 스트림의 캐릭터 워드를 출력하고 프리픽스 워드를 캐릭터 워드로 대체하는 단계와,

코드 스트림의 다음에 위치하는 캐릭터 워드를 읽어서 가져온 상태에서 가져온 캐릭터 워드가 dictionary에 있는가를 확인하는 단계와,

dictionary에 있는 경우에는 캐릭터 스트림으로 가져온 코드 워드를 출력한 후 프리픽스로 대체하는 단계와,

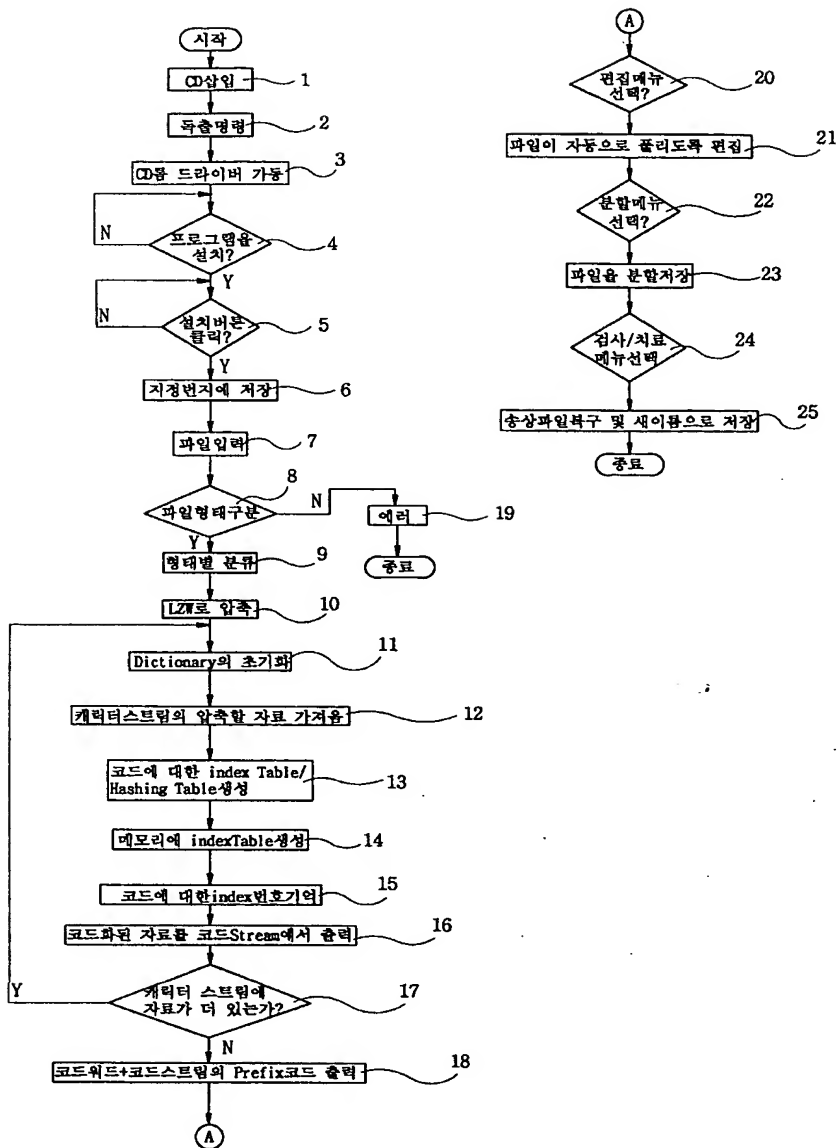
코드워드에서 새로운 데이터를 읽어서 가져온 후 프리픽스 새 데이터를 dictionary에 추가하는 단계와,

상기의 dictionary에 없는 경우에는 프리픽스를 읽어서 가져온 후 코드 스트림을 프리픽스로 대체하는 단계와,

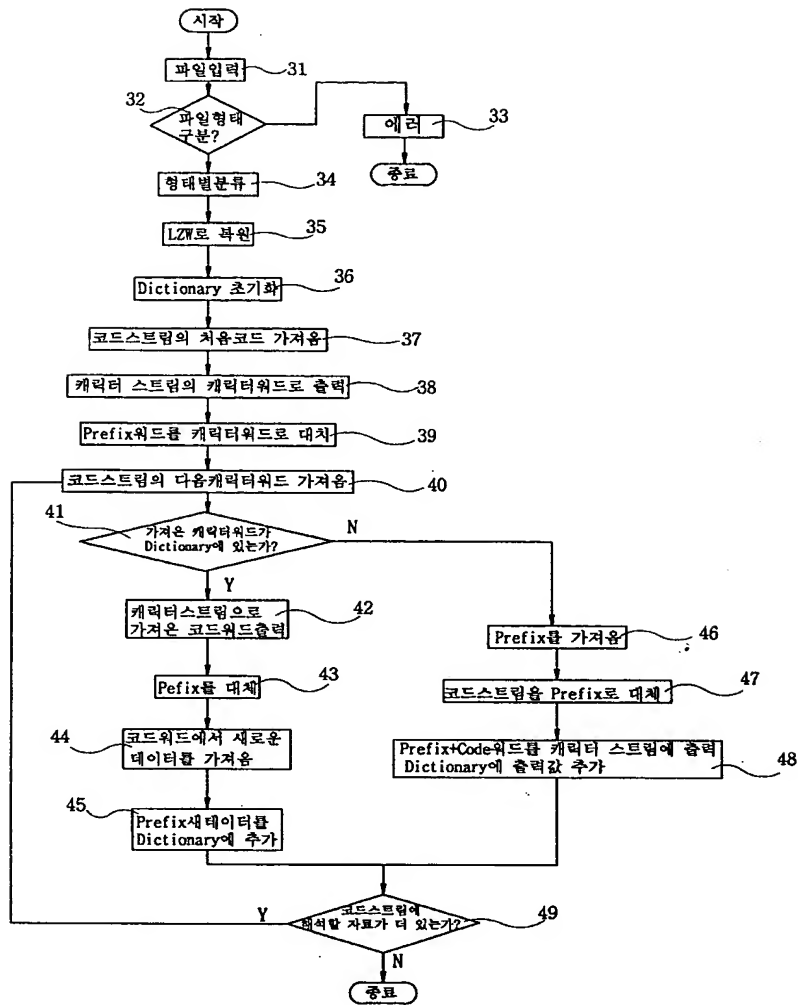
프리픽스 + 코드 워드를 캐릭터 스트림에 출력하면서 dictionary에 출력값을 추가한 후 코드 스트림에 해석할 또 다른 해석할 자료가 더 있는가를 확인하는 단계들에 의해 복원을 수행하도록 한 한글 데이터의 복원방법.

도면

도면1



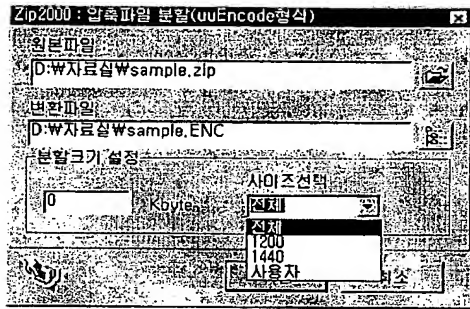
도면2



도면3



도면4



도면5

